Vitro: Designing a Voice Assistant for the Scientific Lab Workplace

Julia Cambre[1*], **Ying Liu**[2], **Rebecca E. Taylor**[2], **Chinmay Kulkarni**[1*]
[1]Human-Computer Interaction Institute, [2]Department of Mechanical Engineering
Carnegie Mellon University
jcambre@cs.cmu.edu, yliu5@andrew.cmu.edu, bex@andrew.cmu.edu, chinmayk@cs.cmu.edu

ABSTRACT
This paper investigates whether voice assistants can play a useful role in the specialized work-life of the knowledge worker (in a biology lab). It is motivated both by promising advances in voice-input technology, and a long-standing vision in the community to augment scientific processes with voice-based agents. Through a reflection on our design process and a limited but fully functional prototype, Vitro, we find that scientists wanted a voice-enabled device that acted not a *lab assistant*, but *lab equipment*. Second, we discovered that such a device would need to be deeply *embedded* in the physical and social space in which it served scientists. Finally, we discovered that scientists preferred a device that supported their practice of "careful deviation" from protocols in their lab work. Through this research, we contribute implications for the design of voice-enabled systems in workplace settings.

Author Keywords
voice assistant; design research; conversational agent; augmented scientific workplace

CCS Concepts
•**Human-centered computing** → **Natural language interfaces;** *Sound-based input / output;*

INTRODUCTION
Recent studies suggest that 46% of adults in the United States use voice assistants like Siri or the Google Home [39], and projections estimate that there will be upwards of 5 billion assistants installed on smartphones worldwide by 2022 [42]. These systems promise boundless convenience, particularly within "hands-busy, eyes-busy" situations [48]: with a simple voice command, users can, in theory, control home lighting systems, check the weather, buy goods, and so on [47, 45, 26].

* These authors contributed equally to this paper

DIS '19, June 24–28, 2019, San Diego, CA, USA.

ACM ISBN 978-1-4503-5850-7/19/06 ...$15.00.
http://dx.doi.org/10.1145/3322276.3322298

At the same time, much of this potential is unrealized: users often treat these devices as little more than exciting toys [47]. This has led some researchers to question whether voice assistants have any compelling utility in everyday life [33].

Against this backdrop, this paper investigates whether voice assistants can play a useful role, not in the everyday life of a household, but in the specialized work-life of the knowledge worker. It is motivated both by promising advances in voice-input technology (e.g. see [6]), and the long-standing vision of the scientific community for such help. Bush's 1945 essay, "As we may think" [7], popularized a future vision in which scientists freely wandered about their labs, taking notes and documenting their observations not on paper, but through a head-mounted camera and a dictation system. In the decades that have followed, it has inspired many projects that use voice to augment the scientific process (e.g. [5, 3, 16, 6]). Now that the technology is within imminent reach, we ask: what might it do, and how?

This paper has modest goals: it does not seek to report on a new AI technology, nor does it purport to introduce a new design technique or methodology. Instead, it simply asks if (and how) human-computer interaction and design practitioners and researchers may take voice as a serious design material for hands-free workplace applications. In doing so, we reflect on our design activities and our designed artifacts, how knowledge workers responded to our design choices, and design opportunities they (and we) then identified.

We present this paper as a case study in the design of Vitro, a voice-based assistant we created for a biology wet lab. Specifically, Vitro helps highly trained biologists with *culturing cells*, an exacting process for keeping cells alive in an artificial environment. Biologists follow *protocols* for cell culture (much like recipes, but requiring far more precision), and Vitro guides them through one key protocol as part of this process for "passaging" cells, or splitting them into new plates.

In designing Vitro, we conducted a contextual inquiry process that involved shadowing lab members for eight months, examining cell culture protocols and common errors in following them, and reflecting on what Vitro should do. Based on this inquiry, we designed Vitro with a limited but fully functional set of abilities: reading out steps, helping with step timing, and providing more information about the protocol on demand. Because biologists would be interacting with Vitro dozens of

times throughout the protocol, we also designed it to be continuously available without needing an invocation command (such as "OK Google..."). Finally, we studied how five lab scientists used Vitro in the process of cell culture.

Through this work, we discovered that first, contrary to our initial assumptions, many scientists wanted not a *lab assistant*, but *lab equipment*. Instead of an assistant with conversationality and flexibility, they preferred equipment that was limited and reliable. Rather than a human-like coach who helped with cell culture, they preferred a system that merely presented statistics on common mistakes. Second, we discovered that such a system would need to be deeply *embedded* in the space in which it served scientists. For instance, instead of an assistant that told them to pipette a solution, scientists wanted a system that told them which specific pipette to use. Third, we discovered that scientists preferred a system that supported their practice of "careful deviation" from protocols in their lab work. Finally, scientists saw Vitro as belonging not to them, but to the lab (consistent with their view of it as lab equipment), suggesting that designers must carefully consider that their user may not be one person, but the many people who work in a space.

BACKGROUND AND RELATED WORK

Augmenting scientific environments

There have been several approaches to scaffolding and supporting procedural tasks in the scientific workplace. Often, these approaches are the result of new design materials such as pen-and-touch [56], sensor-augmented environments [4] or tabletop computing [51]. These earlier works serve as inspiration to our own research, and inform our own contextual inquiry in our deployment setting.

Tabard et al. [51] suggested four requirements for augmenting biology labs (our case setting) in particular: 1) support access to digital resources and services, 2) support capture of experimental data into digital documents, 3) support the iterative experimentation process; and 4) handle the pervasive presence and use of a wide range of heterogeneous artifacts.

Our work departs from this literature in that we are not reticent in questioning these expressed constraints in our target environment, such as the need to support capture of experimental data into digital documents [51, 56]. In our own work, we seek to empathize with scientific needs, but also reframe them to uncover new design opportunities.

Just as with pen-and-touch and tabletop displays, there are many demonstrations of how voice might support a scientific workplace. For example, Austerjost et al. modified laboratory equipment to respond to voice commands, powered by the Amazon Alexa skills kit [6]. Austerjost et al.'s assistant demonstrated how one could take measurements and configure settings on lab equipment (e.g. "set the balance to zero" on a scale) through voice commands. Similarly, Kincaid and Pollock created a voice assistant, Nicky, to answer highly technical, domain-specific questions about test and measurement equipment [25]. Commercial aspirations abound as well: for example, HelixAI, a startup company [1] is developing an app for the Amazon Alexa ecosystem, with the goal of supporting scientists with their lab protocols, ordering and stocking lab materials, and helping with basic calculations, conversions, or definition look-ups [1]. However, while these systems have shown that certain tasks *could* be done, they do not focus on which tasks must be done, and how. In short, they establish technical feasibility; the current paper examines design opportunities.

Finally, previous work has studied the affordances of different emergent media in augmenting scientific work, and how they might be combined. For example, Zheng et al. compared full-coverage mixed reality glasses, peripheral glasses, a tablet, or traditional paper-based instructions for expert task assistance to study how each modality affected users' task performance [57]. Similarly, Funk et al. compared complex task instructions delivered via in-situ projection, a head-mounted display, tablet, or paper, and found performance benefits for reducing completion time and cognitive load with the projection approach [17]. Within the biology lab context, recent research has also explored using augmented reality technology for experimentation support. Scholl et al. designed a custom application for Google Glass to guide users through a DNA extraction protocol in a wet lab, combined with a wrist-worn accelerometer device to automatically infer users' progress through the experiment [46]. In a similar study, Hu et al. conducted a field study of Google Glass and found that using Glass untethered experimenters from physical protocols and provided new affordances for documentation (e.g. through the head-mounted camera); participants found voice input to be convenient, but struggled to interact with the Glass system through gestures and touch-based input [21]. This preceding work is largely complementary to our own. Unlike this work, our focus is not on inventing new media, but in understanding how designers might use the existing affordances of voice assistance creatively in a scientific workplace setting.

Conversational assistance in the workplace

Prior research for conversational assistance in the workplace has studied how the affordances of specifying tasks in natural language (a "conversational" interface) may augment information seeking [10, 29, 28], and reflection [53, 27].

Our research differs from this work in our focus on voice, rather than text. The distinction of a voice versus a text "conversation" may seem nuanced, but is crucial; practices of reading and writing are distinct from oral conversations, which convey meaning through pauses, turn-taking and yielding, and conversational cues [45, 52]. These affordances can qualitatively change the interaction: for instance, Kocielnik et al's work on supporting reflection with both text and voice suggests that voice was seen as more efficient, personal, and engaging, even as text allowed for more considered responses and deeper cognition [27].

Designing with voice

Companies such as Amazon and Google have created developer toolkits and guidelines for practitioners to develop voice based interactions (e.g. [19]). These design guidelines generally suggest how particular conversations may be designed (e.g. suggesting that voice assistants should abide by Grice's

Maxims [20]). However, as a recent literature review by Murad et al. noted, voice interfaces currently lack the sorts of best practices and design principles that have informed the design of GUIs for the past several decades [37]; while heuristics such as recognition over recall and supporting user freedom may transfer, the authors argue that voice interfaces require their own set of design principles because other heuristics and needs may be unique to voice [37]. While these resources offer a starting point for best practices, they do not guide designers about what *workplace actions* might best be supported with voice.

Recent research in how people use voice interfaces suggest opportunities for design. Interactions with voice assistants are both through explicit conversational turn-taking, and implicit conversational cues [52]. Furthermore, users interweave voice commands to devices like the Amazon Echo and the interpersonal conversations around family events (such as dinner) [45]. This calls into question whether voice assistants are "conversational" at all or simply accept commands through voice, while embedded in an *actual* conversation [45]. Unfortunately, given our specific setting (a lab environment where scientists often work alone), we did not plan on examining design issues arising from this embedding. Similarly, because we built our prototype using commercial technology (Actions on Google and Dialogflow), rather than as a Wizard-of-Oz setup we were unable to explore implicit conversational cues.

Research-through-design and reflective practice

We present our work as a case study in a particular biology lab in a North American research university. As a systematic reflection on our design process and our design artifacts, we consider it as an research-through-design process [60]. In so doing, we consider our limited but functional prototype not as a prediction of what future voice interfaces in the workplace *will* look like, but rather as a "proposition" for a preferred state [50].

In this process, we see our work as design-researchers as "bricoleurs" [32], not as inventors. This motivates our choice to use existing commercial technology, albeit in a hopefully creative way. At the same time, we are interested in working with scientists to understand how they might use voice assistants with their limitations in their actual work; we thus consider our research setting as a "field" with all its inconvenient realities [58]. This motivates our decision to create a functional prototype, rather than a Wizard-of-Oz design.

CASE STUDY SITE: CELL CULTURE IN A WET LAB

To study the design of voice assistance in a scientific workplace, we partnered with a small biology research group at our university in the United States. Specifically, we focused our inquiry on the process of *cell culture*.

Performing cell culture is essential to all of this group's work, and to biological experimentation in general. Cell culture refers to the process of growing cells over time in a human-controlled environment such as a Petri dish. The research group studies biomaterials and biomolecules for medicine (specifically, using DNA nanostructures as biosensors and/or drug delivery agents), and culturing cells is a task of fundamental importance. For example, the biomanufacture of bio-materials such as heart tissue relies on culturing stem cells, and exposing these cells to growth factors that lead to cell differentiation into heart tissue (see [9, 35] for a review). Similarly, to find medication that is personalized, cells cultured from an individual may be exposed to particular drugs to measure drug reactions (e.g. [30]). Cell culturing is thus a daily, long-term fixture in biology research groups like the one we studied.

Cell culture lab work is generally performed in a "bio-hood" to keep the cells in a sterilized environment. When in use, bio-hoods produce a substantial amount of noise due to HEPA air filters. Though this research group works in many lab environments, we studied in detail a "wet lab" used for culturing cells. Physically, this lab space is small (about 8ft deep and 16ft wide), and noisy (approximately 68 dB on average throughout the space).

Contextual inquiry

We conducted eight months of interviews, observations, and collaborative design sessions with the biology research group to understand how biologists performed cell culture, and how computational task assistance could augment their current practices. During this time, we worked extensively with three research assistants—RA1, RA2, and RA3 (who is also the second author of this paper)—whose daily responsibilities included performing cell culture to study how they developed expertise in cell culture, and how they transferred that expertise to others. One scientist held primarily responsibility for caring for the cells at any given time (RA1 at the start of our observation), but we witnessed two periods of transition in which one research assistant trained the next in order to take over the process.

The first author maintained a paper diary with shadowing observations, and made particular note when the research assistants stressed the importance of processes and artifacts. She also made notes of her own reflections, both during and between observations. Typically, these reflections tended to focus on apparent incongruities between research assistants' articulated and actual process. Where possible, she followed up on these differences in our next shadowing session, and noted their explanations. These handwritten notes were augmented by pictures of the lab space and setup (when it was safe and unobtrusive to take pictures.) For particularly expertise-laden processes (such as thawing and re-plating cells), we recorded video with a smartphone. We periodically synthesized our observations, and verified them with the biology research team.

On synthesis, our observations led to three recurring themes: *centrality of protocols*, *training through apprenticeship*, and *careful deviation*.

Cell protocols: Crucial sources of truth

Protocols provide a step-by-step explanation of a scientific process. They are intended to be both a statement of intent ("how should these cells be handled?") and a repository of

collective knowledge. The cell culture protocols in the observed lab were printed instructions consisting of a series of steps, with branching instructions based on what technicians observed in their cell cultures. For example, protocols would mention procedures for checking the "confluency" of cells (how densely packed the cells were on the plate) and choosing to either re-plate the cells in new media, split the cells into multiple plates, or wait for the cells to continue growing as-is.

Training through apprenticeship

Training in the lab environment followed an apprenticeship model. The two transition periods that we witnessed in which one research assistant trained another on how to conduct cell culture involved a scaffolded process over three or more sessions. Research assistants were apprenticed regardless of whether they had done cell culture before, and were taught the specific cell protocol the lab used. Apprenticeships starting with shadowing, and moved to culturing with supervision, and culturing with limited supervision. This apprenticeship model was put in place to capture tacit knowledge, and to limit protocol drift (below).

Careful deviation

Lab members described *studying* protocols before starting on a new one; they would look over each step to identify any terms or techniques that were unfamiliar to seek clarification. Others described mentally walking through the procedure to optimize their time by seeing if any steps could be handled in parallel (e.g. labeling new materials during a waiting step). With practice, lab members said, they internalized steps so they rarely (if ever) consulted the paper protocol. This process of study suggested our participants were methodical and *careful*.

Inexact protocols: In theory, protocols spelled out procedures in enough detail that another scientist could replicate it and yield the same results [31]. In practice, however, protocols often fell far short of this ideal. One scientist told us how they once paused mid-experiment to make a phone call to a labmate who had previously "owned" a protocol for clarification.

We also observed a phenomenon which the research assistants called "protocol drift," which is a common occurrence in labs and a frequent reason for non-replicability of results. For example, RA1 had performed cell culture almost daily for several years, and had their process fully internalized. As they transitioned away from the lab, it became apparent that over time RA1 had developed tips, tricks, and modifications that had never been documented because there was no immediate need to do so. As a result, the paper version of the protocol had fallen significantly out of date with current practice.

Inexact practice: When RA1 departed from the lab, the lab had to quickly train another member of the group to care for cells RA1 had previously "owned". It then became clear that *confluency*, despite playing a critical role in determining what action the scientists took on their cells, was "eyeballed" rather than precisely measured. Eyeballing allowed RA1 to develop a tacit knowledge of when cells were overly confluent [44] (and needed to be "passaged," or split into a new flask). However, eyeballed estimates within the group varied, with the PI and the newly hired technician differing by approximately 25% – a range large enough that they might take different actions. As each lab member became proficient with the protocol, they developed their own tacit understanding of what was "confluent enough." Similarly, lab members played around with other parameters of the protocol, such as the speed and time to centrifuge a batch of cells, largely based on their tacit knowledge. Often, these tweaks to the protocol were an effort to troubleshoot when cells failed to grow as expected, or to compensate for under-specified details in the written steps.

Taken together, the inexact protocols and practice lead us to think that scientists engaged in a kind of *careful deviation*. While some decisions may have changed how cells grow in unpredictable ways, the scientists knew most adjustments they made would be harmless.

REFLECTIONS ON LAB ACTIONS

Based on our synthesized observations, we reflect on the practice of cell culture and how it may be augmented with voice assistance. Note that as with all research-through-design, these reflections are not "scientific findings" of the world-as-it-is, but rather proposals of what preferred futures might be [58].

First, given the centrality of cell protocols to lab practice, we hypothesize that voice assistance might successfully focus on protocols too. This, already, is a departure from current approaches, as protocols are not atomic actions like "Stop playing music," but lengthy lists of related actions, none of which are meaningful by themselves. Nor are they simply recipes, but complex branching instructions that rely on scientists' tacit expertise. For instance, protocols may specify different courses of action depending on how long cells have remained in the same confluent state.

Second, the careful sloppiness of lab work suggests that capturing experimental data digitally is neither central, nor perhaps always feasible (at least in our study setting). Most deviations from the protocol are harmless, and requiring scientists to document every difference may be unnecessary. Furthermore, given the degree of tacit expertise and action, it may not even be possible to document this exactly (e.g. consider the case of estimating confluence.) Indeed, voice assistance may leverage the opportunity that some careful deviation may be acceptable, as long as it does not lead to protocol drift.

Third, the complexity of the cell protocols suggests that a purely auditory output would be insufficient. Together, we used these reflections to prototype Vitro, described next.

VITRO: A VOICE-BASED LAB ASSISTANT FOR CELL CULTURE

Given the apprentice model of training we saw in the lab, we created Vitro to be an always-present, task-oriented lab assistant, who would scaffold learning a *particular* protocol just as a labmate would. The prototype went through several design iterations based on feedback from our lab collaborators. We describe the final iteration below.

Vitro provides users with step-by-step directions on a cell protocol (based on cell protocol centrality). Currently, Vitro is designed the support a cell passage protocol in culturing cells

(informally, splitting cells into new plates to provide them with more room to grow). To interact with Vitro, users speak a command to the assistant (e.g. "What's the next step?" / "Can you repeat that?"). The primary mode of *input* is through spoken English. Its output is multimodal: Vitro responds with synthesized speech, and displays supplemental information on a visual display.

To support "careful deviation," we designed Vitro such that participants could interrupt it at any time (including mid-speech), skip to a different step in the protocol, or mark a step as complete.

Given the length and complexity of cell culture protocols, scientists would need to interact with Vitro approximately once a minute. To reduce the friction of the interaction, we designed Vitro to respond to any voice it detected without a wakeword to activate the assistance. This is in contrast to current voice assistants that have typically brief sessions of "request/response" interactions [45], each of which must be started with a "wakeword" (e.g. "OK Google" or "Alexa"). The absence of a wake word or explicit trigger also makes it possible to *interrupt* Vitro's responses and cede control back to the user immediately.

Because current commodity assistants consider the conversation "closed" if they do not receive a response within a brief window (5 seconds for an Amazon Echo device [14], and 8 seconds for the Google Home [15]), we modified the framework we used to capture voice continuously. Vitro's conversational interface is built with Actions on Google, a toolkit for implementing custom applications for the Google Assistant. The Actions on Google platform handles speech recognition and integrates with Dialogflow, which performs the backend logic for the voice assistant. Specifically, Dialogflow maintains a model of the conversational *context*, and works primarily by slot-filling using contextual cues and sample utterances to select an appropriate *action*. Once the *action* is selected, Dialogflow forwards the request to our custom API, which produces the text and supplemental information (e.g. images) for the response. Finally, the response gets forwarded back up the chain to Actions on Google, which produces speech synthesis (TTS) in English with a female-sounding, American accent. We use Google's Firebase product to manage the system's backend.

To be resilient to lab noise, we used Apple AirPod wireless Bluetooth headphones for audio input and output, which paired with a MacBook Pro displaying the Vitro graphical interface at full size on a 13-inch Retina display.

We have open-sourced the code for Vitro for others to modify and reuse at `vitro.expertiseatscale.org`.

Multimodal interaction

Given how noisy the lab was, and the degree of detail in the cell passage protocol, we designed a visual display (Figure 1) to augment spoken instruction. Where possible, we used a sparse display with bold, large text to ensure that users could more easily read the content on screen from a distance, motivated in part by the design of electronic checklists in medical settings [54]. The bulk of the display focuses on the current

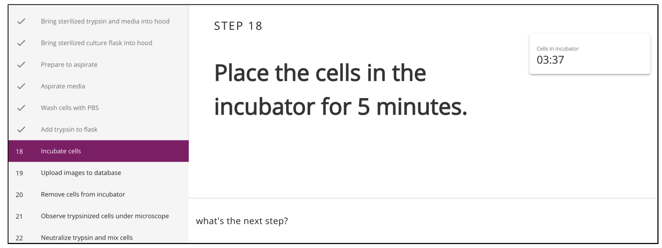

Figure 1. Vitro's visual interface displays the step content, a checklist of the protocol, live transcriptions of the users' speech, and supplemental information like a timer countdown

active step, and shows the full text of the current step, mirroring Vitro's voice responses. Vitro also occasionally displays supplemental information along with the step text, such as images or timers. For example, if the step involves centrifuging cells for 10 minutes, the interface will automatically show a timer with a live countdown.

A sidebar on the lefthand side of the screen displays a checklist-style view of the full protocol in small lettering for context. Additionally, a small panel at the bottom of the interface displays a live-updating transcription of the user's speech. When the microphone is active and detects speech, the background of the transcription panel turns blue to indicate its listening status.

By default, voice assistant toolkits for the Google Assistant and Amazon Alexa provide minimal flexibility for visual interface design components, and instead model all conversations around a chat-like interface with multimedia card elements. To implement Vitro's interface, we extracted the text and audio responses from Actions on Google and presented them through a custom Electron desktop application built with Vue.js and stylized with the Material design framework.

FIELD STUDY

We conducted an exploratory study of Vitro with five scientists who were previously uninvolved with our project. The main objectives of this study were to understand how scientists who had not participated in our design process would use our prototype, probe their mental models of lab voice assistants to understand their beliefs and understanding towards voice assistance (similar to other technology in the past [24, 22, 11]), and to uncover further opportunities for design.

Participants: Five scientists (2 male, 3 female, average age 23) from our university participated in the study. All participants were undergraduate and graduate students affiliated with the biology research group we collaborated with, but had not performed cell culture on the specific type of cell line. Participants gave written, informed consent, and were compensated $20 for their involvement in the study. All participants were required to be familiar with at least the basics of lab technique (e.g. how to properly use a pipette and microscope) and have the mandated Biosafety Level 1 (BSL-1) training certification to use the lab.

Despite their training, participants varied considerably in their familiarity with cell culture, and with the specific procedure involved in this study. On a pre-task survey, one participant

indicated that were not at all familiar with cell culture, and all others indicated they were at least moderately familiar, with a mean prior self-identified familiarity of 2.8 on a 5-point Likert scale. Four participants self-identified as native speakers of English; one self-identified as “moderately proficient” in spoken English.

Study protocol

We began the study by asking participants to complete a brief pre-task survey which asked about their familiarity with cell culture and experience and perceptions of voice assistants and voice-enabled devices. We also notified them that we would be video-recording them in the lab and taking pictures, and that they would be using a voice assistant that sent their voice data to the cloud. We then outfitted participants with the requisite safety gear (a lab coat, goggles, and nitrile gloves), and provided them with a pair of sanitized Apple AirPod wireless Bluetooth headphones to wear during the study (Figure 2).

After a sound check, RA3 provided a brief review of the lab’s equipment and safety protocols. Two facilitators were present in the room at all times: RA3, a trained cell biologist, ensured lab safety and answered any questions that Vitro could not address, and the first author helped with technical issues.

Once in the lab, we explained that Vitro was an experimental system that would guide the participant through the cell culture protocol (specifically, the protocol for passaging cells) step-by-step. We did not provide specific guidance as to what prompts Vitro could understand, or the scope of its understanding, but did say that the assistant was “always listening” for them to speak a command. We also taught them a hand signal they could make to ask us to temporarily disable Vitro’s microphone. A paper version of the same protocol was placed immediately next to the biosafety cabinet (the sterile enclosure for handling cells); participants were told they could reference the paper copy if they wanted. We provided a paper copy to see if participants would simply give up on Vitro at any point. In practice, none did (perhaps because we were present in the room).

Participants used Vitro to follow a cell passage protocol for 3T3 fibroblast cells [2]. They had up to 80 minutes to complete the protocol. As participants interacted with Vitro, we captured video and audio data, logged the transcriptions and interactions through Vitro, and took written notes. After finishing the protocol, we asked participants to complete a post-task survey, followed by a semi-structured interview lasting approximately

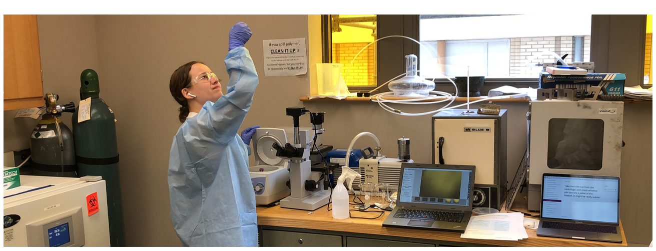

Figure 2. Participants communicate with Vitro through wireless headphones, and can reference protocol steps and supplemental information on a laptop screen (far right)

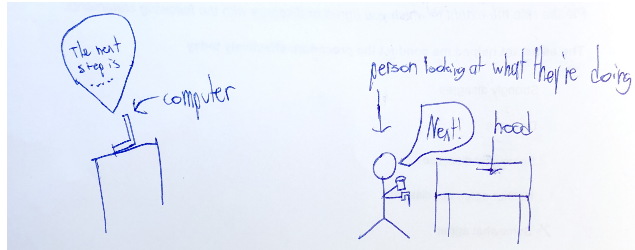

Figure 3. P4’s sketch of how Vitro works

15 minutes. Interviews were audio recorded and transcribed for analysis.

Finally, to elicit participants’ mental models of how Vitro worked, participants were told to *“Draw a quick cartoon sketch of how you think the assistant works, as if you were explaining it to a 5-year-old.”* During the interview, we asked participants to verbally walk us through their sketch to explain their understanding.

FINDINGS

Several themes emerged from our observations of lab sessions and interviews with participants.

A simple piece of lab equipment

Most participants’ mental model representations of Vitro were as flowchart-like diagrams. All drawings and explanations implied that the user operates the system (similar to operating lab equipment), directing Vitro with their voice. Two participants focused on the physical affordances of having a voice-enabled device within the lab environment. P4 described his sketch (Figure 3) as follows:

> This stick figure drawing is saying a person with what appears to be a beaker and with pipettes in their hand, looking at what they’re doing, working in the hood saying “Next” and the computer telling them what the next step is. So, it’s a way if your hands are full, and you’re balancing two open containers or pipettes [...], you really gotta concentrate on that and you can’t really look away [...] You can **just yell "next" and it’ll tell you**, or "repeat that step" or something like that and it’ll tell you. [...] And that I think is **the most useful function of it** because sometimes your eyes have to be focused on what you’re doing and not what you’re *going* to be doing. (P4) [emphasis ours]

When participants described dependence on the assistant, they similarly tended to credit machine-like attributes such as completeness and reliability, rather than personal attributes like conversationality or flexibility:

> The first time it started, maybe it was me not knowing that system well enough. I think three steps down, I felt very... very in the zone. [...] I feel like I could just depend on it to tell me the next steps. I think that took a level of trust in the system [...] Like when I first walked in I was just like, “well, I don’t know if I can trust it.” [... Only after a couple of steps] I realized that [the spoken

> protocol] was fairly concrete, it was fairly descriptive that then I could let my guard down. (P5)

In sum, these observations suggest to us that participants saw Vitro as lab equipment, rather than as a labmate. This observation comes with two caveats. First, not all participants saw Vitro as equipment. With other participants, the boundary between person and equipment seemed blurry. For instance, P5 explicitly described Vitro as "an android" and described the process as "like having a second person watch over you, but it was not really a person, it was just technologically bound."

Second, participants' perceptions of Vitro may have been influenced by our own actions. In particular, the name Vitro is not a common human name (unlike e.g. Alexa), which may have suggested a non-anthropomorphic assistant. Further, in our experimental protocol, we were careful to refer to Vitro by name or as "the system" rather than through personal pronouns (e.g. he, she, they).

Overall, participants did not seem either positive or negative about their lab-equipment metaphors. Whether or not participants saw Vitro as equipment on their own, or because of how we as the designers referred to it, this equipment framing does open up new design opportunities.

Designing for careful deviation

Despite our initial observation that scientists engaged in "careful deviation" from protocols, Vitro supported this practice only minimally. Our participants' reactions seem to reiterate this need, albeit couched as personalization and gathering statistics.

For example, one participant noted how shorthanding directions might help, and suggested that users should be able to pick important checkpoints, in effect allowing them to decide what points in the protocol require care, and where deviation or reduced guidance is appropriate:

> [People who do these procedures every day] don't really need very descriptive points but can be like "check trypsin, check this." [...] It could just be a two word statement that you can skip through [...] or you just have reduced checkpoints along the way. [...] The other thing would be if you can do something to the effect of, "Here's the 35 point protocol. Pick and choose the ones you would need checkpoints for." You essentially tell it to skip steps beforehand. (P5)

Similarly, P2 and P5 wanted statistics which might help them better identify points where care was essential.

> It would be so cool if it was learning from all the people who were using the system. Like "95% of users get this step wrong" (P2)

> It's nice to have something that always says, "be very careful because it's one of these steps that people constantly miss." (P5)

While most participants suggested changes that would better support strategic deviation, P1, P2, and P5 also reflected how it might help them become more strategically careful:

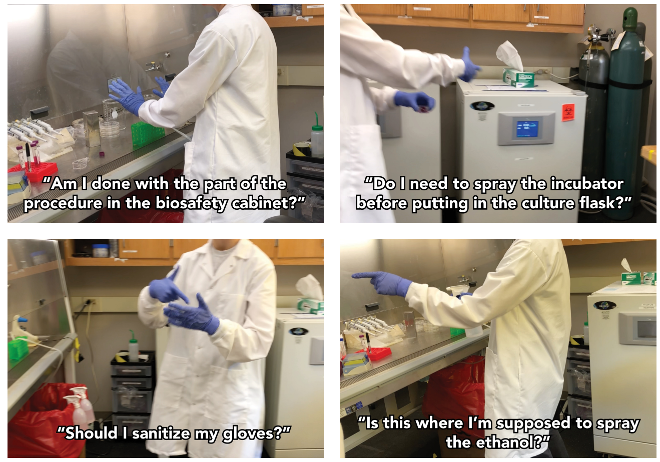

Figure 4. P2 gesturing to silently communicate with the expert facilitator without Vitro's intervention. The text on each picture indicates the question implied by the gesture.

> If you were able to talk to Vitro and be like [...] "Note: cells have fully detached, appear rounded," things like that, that would be something we'd have to type up in our lab notebooks. (P1)

Embeddedness in space

For all but one of the participants in the study (P5), the voice-based steps under-specified many crucial details about the physical lab space. For example, a step that calls for the participant to "pipette the solution" was ambiguous because there were three types of pipettes available in the hood. Implicitly, all participants assumed that because Vitro was available in a lab space, it would be fully aware of the lab equipment and layout. Several participants expected Vitro to offer help with specific lab equipment they were unfamiliar with:

> This is because I'm a novice, but if it had [...] a heads up if you're about to enter the sterile environment, so a cue that's like "you need to wipe your hands off before you do this step" and then another cue to say things like "Use the motorized pipette to do it, and do you want a quick video on how to use it?" (P2)

The pitfalls of wake-word free interaction

In the design of our prototype, based on our lab observations, we spent substantial engineering effort on enabling a wake-word free interaction. Unfortunately, it was almost uniformly a poor design decision. There were several occasions in which participants forgot the microphone was on, and Vitro unsuccessfully attempted to parse a request that was meant only for the other human research assistants in the room. Other participants pantomimed gestures and mouthed questions to facilitators so the system wouldn't hear them (Figure 4).

These observations suggest that whatever the putative interactional benefits that wakeword-free activation may have, they will require the system to achieve *attentional awareness*: when a user is primarily following steps through voice commands, but may also need to communicate with others, the assistant should be able to infer whether utterances are directed towards

it. We also overlooked that participants might want to talk to themselves; for example, P2 described how they might like to engage in auditory rehearsal [52]:

> I was thinking when I was doing the 20 pipettes to break up the cell pellet [*one of the protocol steps*], that if I could say out loud, '1, 2, 3...' and then if someone came over and started talking to me I could look back and say 'Oh I was on 4,' and then say, 'Ok 4, 5, 6, 7...' (P2)

These findings, combined with our observation that voice-based systems may be seen as equipment, rather than assistants capable of fluent conversation [8], suggest that designers may want to consider a low-effort activation mechanism instead.

Technical issues with voice interaction

Transcription errors were the most common issue that participants encountered, and the most frustrating. This may be due in part to the technical vocabulary (e.g. "confluency" or "trypsinize") required as part of the cell passage protocol, consistent with findings from contexts such as song and artist name recognition [49]. Participants frequently compensated by hyperarticulating, raising their voice, or moving closer to the screen displaying Vitro's visual output; these strategies have also been observed in prior work on voice user interfaces [45, 38, 40].

Prior work has also found that voice assistants generally suffer from low discoverability [55, 37]. Vitro was no different; participants rarely explored Vitro's capabilities:

> I didn't know the full range of things I could ask it. So in my mind it was very much a "oh, they're just going to read me off the protocol" [...] I think it would've helped to know "these are the top 3 to 5 things you can do with it" (P1)

Privacy and possession

On average, participants in our experiment were only "moderately comfortable" with how voice-enabled devices such as Google Assistant, Siri or Cortana handle user privacy. Three participants reported never using voice-based assistants; the remaining two participants used them "once a week" and "several times a day."

Previous work by Cowan et al. has found that one of the primary reasons why people choose *not* to use commercial voice assistants regularly is a concern over privacy, with specific concerns about the voice assistant's connection to sensitive personal data kept on smartphones and a fear of misuse, such as financial information, health data, and personal contact information [12]. The infrequent users in Cowan et al.'s study also reported concerns that the assistants were "always listening" or otherwise storing and sharing their data for profit [12].

However, while participants may not have trusted current commercial assistants, they expressed relatively high levels of trust and comfort with Vitro; the average self-reported trust in Vitro was 5.4 ("somewhat agree") on a 7-point Likert scale. This is likely because Vitro seemed simple relative to other voice assistants:

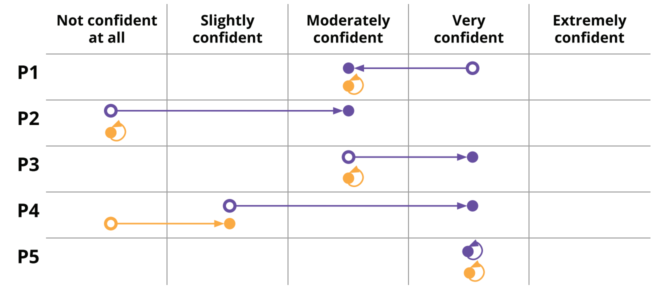

Figure 5. Conducting cell culture with Vitro helped improve participants' understanding (in purple), but had minimal effects on other measures like detecting issues with the cell culture process (in orange). Open circles represent pre-task responses; closed circles represent post-task responses.

> I feel like in terms of Alexa and Siri, they just have a lot stored in them, whether through the cloud or through their hardware. Whereas I wasn't exactly sure how much was stored in here, and it's probably just a very small percentage of what something like [Alexa and Siri] would have the capacity for. (P1)

Possession: Participants frequently suggested features that indicated that even while the assistant helped them personally, it did not belong to them, had little access to their personal information, and could use their interaction data in positive ways to help others in the workplace. P2 elaborates:

> I think the difference, whether it's true or not, mentally what I'm thinking is, here [in the lab, with Vitro], I'm not providing a lot of personal information, but all that crap on my phone... you know, there's bank accounts and stuff, and that worries me. But if it was in a lab, and was just like, "How many cell cultures did [name] screw up today," I'd be fine with that. I much more trust it in the work setting than in personal places. (P2)

We thus believe there may be a design opportunity to create domain-specific assistants which have only limited information to allay general privacy concerns and design for the work group, rather than the worker. At the same time, we note that a device controlled by an employer may also introduce new privacy and trust considerations. In our study, the strong Institutional Review Board-backed guarantees of privacy and consent may have precluded such concerns.

Other findings

Participants in our study reported higher confidence in understanding the steps involved in cell culture after using Vitro; self-rated confidence scores increased by an average of 0.8 points on a 5-point Likert scale between a pre and post-task survey (Figure 5). Other self-rated confidence measures, such as ability to perform the steps involved, detect problems (Figure 5), and replicate the protocol remained largely the same.

Finally, all participants were able to complete the cell passage protocol with Vitro's support within the allotted 80 minute timeframe, and successfully transferred cells to a new flask as measured by cell growth over a 1 to 2 day incubation period. However, we do not present this as evidence of Vitro's

effectiveness, as RA3 was present to monitor lab sessions, and because cells may sometimes passage successfully even if the protocol is imperfectly followed.

DISCUSSION

Workplace settings such as the biology wet lab represent a different context for voice assistants than the home. Here, we reflect on our findings and our design process and suggest implications for future design work.

Studying voice in the Field

Design studies of new voice interactions have generally relied on Wizard-of-Oz prototyping (e.g. see [52, 36]), in effect treating the setting either as a "lab" to test some predicted interactive capabilities such that interactions are meticulously perfect (or imperfect in controlled ways), or as a "showroom", where researchers (and participants) engage in critical design [59]. Of course, such work has enormous benefits. However, what such an approach cannot do is engage participants in seeing a proposed future in all its complexity, including the ways in which proposed designs may fail unexpectedly. To our knowledge, this is the first design study of voice interactions in the field with limited yet functional prototypes.

One result of such an approach is in revealing user intentions and metaphors [18, 43]. In our work, this resulted in participants revealing their models of the assistant as not a person, but as lab equipment. This is in contrast to previous work that has suggested that voice assistants (can and) should act as a "butler" [41] in the home. We do not believe that our research contradicts this; rather, it suggests that what might be desirable in the home may be well be different from the lab or office. Similarly, participants' comfort with sharing information with other scientists in the lab suggests that new metaphors of possession may be important in work settings.

The challenge of embeddedness

As designers of Vitro, we unconsciously thought we were designing an electronic labmate, similar to previous work that tries to create an electronic butler. From this point of view, expectations of embeddedness are somewhat surprising. After all, scientists work in multiple labs, many of which may be unfamiliar. We expect many designers of voice assistants have a similar implicit view of voice assistants as human-like, with character and personality [23].

On the other hand, if voice assistants are equipment that is part of the lab, an expectation of embeddedness is easier to understand. While we note in our observations that participants expected *some* level of embeddedness, more study is needed to understand these expectations in detail. For instance, might participants expect the assistant to know if supplies for reagents were running low? What about knowing if other equipment was broken? Addressing these expectations intelligently will be an important design challenge.

Finally, it is unclear if voice is the best modality for helping *participants* themselves become embedded in the social and physical space [34]. As a result, designers may need to weigh the trade-off between making voice assistance equipment more deeply embedded in space while still allowing participants to understand and use the space effectively.

Supporting careful deviation practices and multiple users

While Vitro was designed to allow some deviations from protocol (skipping steps, marking steps as complete), we initially believed that deciding where to deviate and where to be careful was purely a scientific decision. However, our participants' responses suggest that what kind of protocol deviations were acceptable is not only *scientifically* determined, but *socially* determined as well. In particular, participants' request for statistics of how other users perform a step (and what their failure rate was) suggests a design opportunity to shape practice using social information.

To use such social design, voice assistants may need to be designed explicitly with the work group in mind. Of course, we are not the first to suggest that designing for groups is important for ubiquitous computing [13]. This prior work suggests a principled approach to doing so, but designers may need to adopt techniques to a work setting rather than the home.

LIMITATIONS

This paper sets out to propose a possible future of voice assistance in the lab. It is by no means the ideal future, or even the only future. As with all field-based design research, our observations and reflections are our own, and other researchers may draw other, equally valid conclusions with our same research process [59]. Furthermore, we designed the system with our biology collaborators, and only studied five participants in detail during a single session of use. As such, our studies are designed for rich qualitative accounts, not for statistically significant results. Finally, while we study work in a biology lab as an example of workplace support, knowledge work and knowledge workers vary widely. Still, we hope this case study will inform future research on how voice assistants may support other workplaces, and in other contexts that involve procedural tasks.

CONCLUSION

This paper is a reflection of the design process and the design artifact of a voice assistant in the context of a scientific lab workplace. Our exploratory study suggests that scientists desire voice assistance that offers equipment-like functionality rather than human-like support, that is able to support a practice of "careful deviation" that is both scientifically and socially determined, and that the assistant be embedded within its physical and social space. We acknowledge that these findings are proposals of preferred futures, and there may be many others. We also acknowledge that building such assistants as we propose may be difficult, but our contribution is that without these proposed designs, voice assistance is unlikely to work meaningfully. Our next step is to build and evaluate an assistant that does implement these ideas, and evaluate it. An important goal in sharing our findings is to encourage a conversation around what assistants in the workplace must do to support workers and scientists, and how they might do so.

ACKNOWLEDGEMENTS
We are thankful to our collaborators in the Microsystems & MechanoBiology Lab who helped this project with endless patience, feedback and generosity, especially Dilhara Jayarathna, and Alexandra J. Eggleston. We also thank the participants who participated in our exploratory study, and Ben Stein for his contributions to Vitro's technical development. All human-subjects research conducted was approved by the Institutional Review Board at Carnegie Mellon University under protocol 2018_00000144. The research reported here was supported in part by the Carnegie Mellon Manufacturing Futures Initiative, and by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the funders.

REFERENCES
[1] 2019a. HelixAI. (2019). https://www.askhelix.io/

[2] 2019b. NIH/3T3 (ATCC® CRL-1658™). (2019). https://www.atcc.org/products/all/CRL-1658.aspx

[3] Gregory D Abowd and Elizabeth D Mynatt. 2000. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 29–58.

[4] Larry Arnstein, Chia Yang Hung, Robert Franza, Qing Hong Zhou, Gaetano Borriello, Sunny Consolvo, and Jing Su. 2002. Labscape: A smart environment for the cell biology laboratory. *IEEE Pervasive Computing* 1, 3 (2002), 13–21. DOI: http://dx.doi.org/10.1109/MPRV.2002.1037717

[5] Barry Arons. 1991. Hyperspeech: Navigating in speech-only hypermedia. In *Proceedings of the third annual ACM conference on Hypertext*. ACM, 133–146.

[6] Jonas Austerjost, Marc Porr, Noah Riedel, Dominik Geier, Thomas Becker, Thomas Scheper, Daniel Marquard, Patrick Lindner, and Sascha Beutel. 2018. Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments. *SLAS Technology* (2018), 1–7. DOI: http://dx.doi.org/10.1177/2472630318788040

[7] Vannevar Bush. 1945. As we may think. *The Atlantic Monthly* 176, 1 (1945), 101–108.

[8] Graham Button, Jeff Coulter, John RE Lee, and Wes Sharrock. 1995. Computers, minds, and conduct. (1995).

[9] Patrick Cahan and George Q Daley. 2013. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature reviews Molecular cell biology* 14, 6 (2013), 357.

[10] Praveen Chandar, Yasaman Khazaeni, Matthew Davis, Michael Muller, Marco Crasso, Q Vera Liao, N Sadat Shami, and Werner Geyer. 2017. Leveraging conversational systems to assists new hires during onboarding. In *IFIP Conference on Human-Computer Interaction*. Springer, 381–391.

[11] Janghee Cho and Janghee. 2018. Mental Models and Home Virtual Assistants (HVAs). *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), SRC05:1–SRC05:6. DOI: http://dx.doi.org/10.1145/3170427.3180286

[12] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, 43:1–43:12. DOI: http://dx.doi.org/10.1145/3098279.3098539

[13] Scott Davidoff, Min Kyung Lee, Charles Yiu, John Zimmerman, and Anind K Dey. 2006. Principles of smart home control. In *International conference on ubiquitous computing*. Springer, 19–34.

[14] Megan Rose Dickey. 2018. Now you can have a conversation with Alexa without screaming 'Hey, Alexa' for every request. (3 2018). https://techcrunch.com/2018/03/10/now-you-can-have-a-conversation-with-alexa-without-screaming-hey-alexa-for-every-request/

[15] Stefan Etienne. 2018. Continued conversation is arriving for Google Home users today. (6 2018). https://www.theverge.com/2018/6/21/17486062/google-assistant-continued-conversations-home-max-mini-us-launch

[16] Gerhard Fischer. 1993. Beyond human computer interaction: Designing useful and usable computational environments. *People and Computers* (1993), 17–17.

[17] Markus Funk, Thomas Kosch, and Albrecht Schmidt. 2016. Interactive Worker Assistance: Comparing the Effects of In-situ Projection, Head-mounted Displays, Tablet, and Paper Instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 934–939. DOI: http://dx.doi.org/10.1145/2971648.2971706

[18] Verena Giller, Manfred Tscheligi, Reinhard Sefelin, Anu Mäkelä, Aapo Puskala, Kristiina Karvonen, and Victor Joseph. 1999. Maypole highlights: Image makers. *Interactions* 6, 6 (1999), 12–15.

[19] Google. 2019. Conversation design. (2019). https://designguidelines.withgoogle.com/conversation/

[20] H Paul Grice. 1975. Logic and conversation. *1975* (1975), 41–58.

[21] Grace Hu, Lily Chen, Johanna Okerlund, and Orit Shaer. 2015. Exploring the Use of Google Glass in Wet Laboratories. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15* (2015), 2103–2108. DOI:http://dx.doi.org/10.1145/2702613.2732794

[22] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My data just goes everywhere": User mental models of the internet and implications for privacy and security. *Symposium on Usable Privacy and Security (SOUPS) 2015* (2015), 39–52.

[23] Rabia Khan and Antonella De Angeli. 2007. Mapping the demographics of virtual humans. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*. BCS Learning & Development Ltd., 149–152.

[24] Sara Kiesler and Jennifer Goetz. 2002. Mental Models and Cooperation with Robotic Assistants. *CHI'02 extended abstracts on Human factors in computing systems* (2002), 576–577. DOI: http://dx.doi.org/10.1145/506443.506491

[25] Robert Kincaid and Graham Pollock. 2017. Nicky: Toward a Virtual Assistant for Test and Measurement Instrument Recommendations. In *Semantic Computing (ICSC), 2017 IEEE 11th International Conference on*. IEEE, 196–203.

[26] Sara Kleinberg. 2018. 5 ways voice assistance is shaping consumer behavior. *Thinking with Google* (2018). https://www.thinkwithgoogle.com/consumer-insights/voice-assistance-consumer-experience/

[27] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. *Proceedings of the 2018 Designing Interactive Systems Conference* (2018), 881–894. DOI: http://dx.doi.org/10.1145/3196709.3196784

[28] Q Vera Liao, Matthew Davis, Werner Geyer, Michael Muller, and N Sadat Shami. 2016. What Can You Do? Studying Social-Agent Orientation and Agent Proactive Interactions with an Agent for Employees. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16* (2016), 264–275. DOI: http://dx.doi.org/10.1145/2901790.2901842

[29] Q. Vera Liao, Werner Geyer, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, and N. Sadat Shami. 2018. All Work and No Play? *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (2018), 1–13. DOI:http://dx.doi.org/10.1145/3173574.3173577

[30] Christoph Lipps, Franziska Klein, Tom Wahlicht, Virginia Seiffert, Milada Butueva, Jeannette Zauers, Theresa Truschel, Martin Luckner, Mario Köster, Roderick MacLeod, Jörn Pezoldt, Jochen Hühn, Qinggong Yuan, Peter Paul Müller, Henning Kempf, Robert Zweigerdt, Oliver Dittrich-Breiholz, Thomas Pufe, Rainer Beckmann, Wolf Drescher, Jose Riancho, Carolina Sañudo, Thomas Korff, Bertram Opalka, Vera Rebmann, Joachim R Göthert, Paula M Alves, Michael Ott, Roland Schucht, Hansjörg Hauser, Dagmar Wirth, and Tobias May. 2018. Expansion of functional personalized cells with specific transgene combinations. *Nature Communications* 9, 1 (2018), 994. DOI: http://dx.doi.org/10.1038/s41467-018-03408-4

[31] Gordon J Lithgow, Monica Driscoll, and Patrick Phillips. 2017. A long journey to reproducible results. *Nature News* 548, 7668 (2017), 387.

[32] Panagiotis Louridas. 1999. Design as bricolage: anthropology meets design thinking. *Design Studies* 20, 6 (1999), 517–535.

[33] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (2016), 5286–5297. DOI: http://dx.doi.org/10.1145/2858036.2858288

[34] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 580–628.

[35] Carlos D Luzzani and Santiago G Miriuka. 2017. Pluripotent stem cells as a robust source of mesenchymal stem cells. *Stem Cell Reviews and Reports* 13, 1 (2017), 68–78.

[36] Moira McGregor and John C Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2208–2220. DOI: http://dx.doi.org/10.1145/2998181.2998335

[37] Christine Murad, Cosmin Munteanu, Leigh Clark, and Benjamin R Cowan. 2018. Design Guidelines for Hands-free Speech Interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18)*. ACM, New York, NY, USA, 269–276. DOI:http://dx.doi.org/10.1145/3236112.3236149

[38] Chelsea Myers, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. (2018), 1–7. DOI: http://dx.doi.org/10.1145/3173574.3173580

[39] Kenneth Olmstead. 2017. *Nearly half of Americans use digital voice assistants, mostly on their smartphones*. Technical Report. https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

[40] Sharon Oviatt, Colin Swindells, and Alex Arthur. 2008. Implicit user-adaptive system engagement in speech and pen interfaces. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* 4 (2008), 969. DOI: http://dx.doi.org/10.1145/1357054.1357204

[41] Sabine Payr. 2013. Virtual butlers and real people: styles and practices in long-term use of a companion. In *Your Virtual Butler*. Springer, 134–178.

[42] Sara Perez. 2017. Voice-enabled smart speakers to reach 55% of U.S. households by 2022, says report. *TechCrunch* (2017). https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report/

[43] James Pierce and Eric Paulos. 2015. Making multiple uses of the obscura 1C digital camera: reflecting on the design, production, packaging and distribution of a counterfunctional device. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2103–2112.

[44] Michael Polanyi. 2009. *The tacit dimension*. University of Chicago press.

[45] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (2018), 1–12. DOI: http://dx.doi.org/doi.org/10.1145/3173574.3174214

[46] Philipp M Scholl, Embedded Systems, Matthias Wille, Unit Human Factors, Health Dortmund, Kristof Van Laerhoven, and Embedded Systems. 2015. Wearables in the Wet Lab: A Laboratory System for Capturing and Guiding Experiments. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), 589–599. DOI: http://dx.doi.org/10.1145/2750858.2807547

[47] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18* (2018), 857–868. DOI: http://dx.doi.org/10.1145/3196709.3196772

[48] Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (2000), 63–65. DOI: http://dx.doi.org/10.1145/348941.348990

[49] Aaron Springer and Henriette Cramer. 2018. "Play PRBLMS": Identifying and Correcting Less Accessible Content in Voice Interfaces. (2018), 1–13. DOI: http://dx.doi.org/10.1145/3173574.3173870

[50] Cal Swann. 2002. Action research and the practice of design. *Design issues* 18, 1 (2002), 49–61.

[51] Aurélien Tabard, Juan-David Hincapié-Ramos, Morten Esbensen, and Jakob E Bardram. 2011. The eLabBench: an interactive tabletop system for the biology laboratory. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11* (2011), 202–211. DOI: http://dx.doi.org/10.1145/2076354.2076391

[52] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 208:1–208:7. DOI: http://dx.doi.org/10.1145/3173574.3173782

[53] Alex C. Williams, Harmanpreet Kaur, Gloria Mark, Anne L. Thompson, Shamsi Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment using Conversational Intelligence. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (2018). DOI: http://dx.doi.org/10.1145/3025453.3026044

[54] Leslie Wu, Jesse Cirimele, Kristen Leach, Stuart Card, Larry Chu, T Kyle Harrison, and Scott R Klemmer. 2014. Supporting crisis response with dynamic procedure aids. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 315–324.

[55] N Yankelovich, G-A Levow, and Matt Marx. 1995. Designing SpeechActs: Issues in Speech User Interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), 369–376. DOI: http://dx.doi.org/10.1145/223904.223952

[56] R Yeh, Chunyuan Liao, and S Klemmer. 2006. ButterflyNet: a mobile capture and access system for field biology research. *Proc. of the SIGCHI conference on Human Factors in Computing Systems* (2006), 571–580. DOI: http://dx.doi.org/10.1145/1124772.1124859

[57] Xianjun Sam Zheng, Cedric Foucault, Patrik da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2125–2134. DOI: http://dx.doi.org/10.1145/2702123.2702305

[58] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer, 167–189.

[59] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (2007), 493. DOI: http://dx.doi.org/10.1145/1240624.1240704

[60] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. 2010. An analysis and critique of Research through Design: towards a formalization of a research approach. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. ACM, 310–319.